Session No. 750

The Implementation of Business Decision-Making Tools in Incident Rate Prediction

Samuel A. Oyewole, Ph.D.
Department of Energy and Mineral Engineering
The Pennsylvania State University
University Park, PA

Introduction

This paper provides an overview of the application of forecasting technique in the prediction of incident rates from a business-oriented perspective. This research shows that traditional business and financial management tools could be used by occupational health and safety professionals in the quest for making realistic advance predictions of incident rates. From earlier research, actual incident rate data were obtained, and new incident rates were predicted, using the double exponential smoothing method. The model validation was achieved by comparing the actual and the predicted incident rates in a one-year period. Forecast accuracy indicated 71.58% with a tracking signal of -4.08. The ability to predict incident rates does not only help in reducing safety intervention costs, but would make it possible for managers and top-level executives to better understand the financial implications and consequences of ineffective safety programs and policies.

Due to the increasing need for safety professionals to effectively present their proposals and plans to the non-safety-oriented top management and financial executives, it is has become highly necessary to incorporate some business decision-making tools into safety and health programs. To set a foundation for the implementation of these business tools in safety and health, this paper uses business decision-making forecasting technique to predict incident rates. Forecasting could be described as the process of utilizing previous or current situations to estimate or predict future unknown situations. Forecasting is often applied to several aspects of human lives (Armstrong, 2001). This predictive tool is often used by many organizations to make strategic (long-term), tactical (medium-range) and operational (short-term) decisions (Chopra and Meindl, 2006).

Forecasting is widely used in meteorology (to predict climate, weather, hurricanes, tornadoes and flooding); supply chain (to make demand forecasts); economics and business (to make predictions for investment on stocks and bonds); transportation planning (to predict the number of passengers); seismology (to predict earthquakes); and petroleum engineering and geology (to predict oil reserves). Some other recent applications of forecasting include population growth, healthcare, sports (to predict winning chances of players and teams), as well as politics (to predict winners in elections). Several decision-makers use forecasting as the readily available tool to plan ahead and make predictions based on uncertainty (Armstrong, 2001). Since a forecast is an inference of what is likely to occur in the future, forecasts do not always provide the accurate estimates of the actual situations; therefore, the accuracy of a forecast is determined based on the differences in the actual and the predicted values (forecast error). It should be noted that long-term forecasts are usually less accurate than short-term forecasts. This is due to the larger value of the standard deviation of error relative to the mean of the data samples or observations (Chopra and Meindl, 2006). Forecasting is used in situations involving time series or trends. In non-stationary (moving) data, forecasting is used to estimate the mean of the probability of the distributions.

Literature Review

Previous research studies have advocated the incorporation of probabilistic processes used in financial decision-making to the risk assessment concepts and methodologies used by health and safety professionals. Toffel and Birkner (2002) argued that, since health and safety programs often require the allocation of financial resources that needs to be approved by business and governmental managers, the utilization of business-oriented techniques into the safety and health policies would make it easier for the management to better understand and appreciate the safety and health programs. Safety and health programs that are designed based on strategies used in business would be easily approved, since the decision-makers and managers are familiar with the financial decision-making and risk assessment concepts.

In their research, Toffel and Birkner (2002) demonstrated the use of incident probabilities, historic outcome information, and incremental impact analysis in the estimation of risks involving multiple alternatives in the chemical process industry. The findings of their research indicated that certain, easily understood and applied probabilistic risk assessment methods used by business planners and managers to assess financial and outcome risks could be adopted into health and safety analysis. The researchers suggested that safety and health programs could be improved upon by linking the business decision-making activities with health and safety risk assessment processes to securing resources. Also, safety and health policies could be easily understood by managers when additional set of tools for health and safety risk assessment are provided. Toffel and Birkner concluded their research by suggesting that the incorporation of financial tools into the evaluation of safety and health programs would make it easy for the safety personnel to consider multiple risks and provide different decision-making alternatives to management.

In agreement with the suggestions proposed by Toffel and Birkner (2002), Iyer et al. (2005) used forecasting techniques to predict incident rates based on the mathematical model developed in earlier research (Iyer et al., 2004). In their research, the values obtained from the forecast were used to validate their model over a period of 22 weeks. Iyer et al. (2005) adopted weighted moving averages and exponential smoothing techniques to identify changes in the statistical relationship between interventions and incident rates (Haight et al., 2001a and b). In their study, the researchers integrated and related past safety performance (incident rates) with the current rates to obtain an estimate of the future incident rates.

The exponential moving average and the moving average of errors were used to obtain the incident rate forecast. Although Iyer et al. (2005) provided the background for the incorporation of forecasting techniques into the prediction of incident rates, the study did not propose any additional study for the investigation of the behaviors of the observed trend. The exponential smoothing method used in the research assumed constant-level time series. This is actually not the case; in reality, incident rates are not constant, and several factors could account for the changes in the trend levels of varying incident rates. Contrary to the forecasting methodologies adopted by Iyer et al. (2005), this study utilizes the trend-corrected exponential smoothing technique (Holt's model) to predict the incident rates based on the developed safety intervention model. Holt's model is an improvement or modification of the simple exponential smoothing method, which accounts for changes in continuing trends (Chopra and Meindl, 2006). Holt's model is very applicable for predicting incident rates due to its trend correction capability. Holt's model is also known as the double exponential smoothing method.

The Double Exponential Smoothing Method

Time series could be described as a set of observations, which are generated sequentially over time. In situations where the set is continuous, then a trend is generated. Double exponential smoothing (Holt's model) is a refinement of the simple exponential smoothing method, but the addition of the trend-corrected component to exponential smoothing takes into account the behavior of any trend in the data. The simple exponential smoothing method works best with data with no trend or seasonality. Simple exponential smoothing does not often make reasonable forecasts in situations where the data exhibits either an increasing or decreasing trend over time. The double exponential smoothing method is therefore designed to address this type of trend situation in data series (Chopra and Meindl, 2006).

The double exponential smoothing method is used for modeling the behavior of time series (trends) based on a simple linear regression equation obtained in situations where the intercept and slope of the plots of the observations vary slowly over a given time. Unlike the simple exponential smoothing method, Holt's model applies unequal weighting on the exponentially decaying parameters so that newer observations get a higher weighting than older ones (Bowerman and O'Connell, 1993). The degree of exponential decay is determined by the parameter α, where $\alpha \in [0, 1)$. The evolving regression equation could be used to make incident rate predictions.

<u>Application of Holt's Model to Incident Rate Prediction</u>

Incident rate forecasts could be obtained using Holt's model when the incident rate is assumed to have a level and a trend (where incident rate (y) = level + trend). It is therefore necessary to obtain the initial estimate of the level (l_0) and trend (b_0) by running a linear regression between incident rate (Y_t) and the time period t as shown in Equation (1):

$$Y_t = m(t) + c \tag{1}$$

The constant (c) measures the estimate of the incident rate at period t = 0, which is the required estimate of the initial level (l_0). The slope (m) measures the rate of change in the incident rate per period, which is the required estimate of the initial trend (b_0). The trend at time (t) could therefore be given as (b_t) while the level at time (t) is given as (l_t), then the incident rate could be predicted using the following trend-corrected exponential smoothing as shown in Equations (2) and (3):

$$l_t = \alpha y_t + (1-\alpha)[l_{t-1} + b_{t-1}] \tag{2}$$

$$b_t = \gamma[l_t - l_{t-1}] + (1-\gamma)b_{t-1} \tag{3}$$

Where α and γ are smoothing constants between 0 and 1.

The initial values l_0 and b_0 could be obtained based on the estimates of the previous incident rates. Summing up Equations 2 and 3, and the τ-step-ahead incident rate forecast from t for $y_{t+\tau}$, the predicted incident rate [$F_{t+\tau}$], at time t is = t + , and could be given as shown in Equation (4):

$$\hat{y}_{t+\tau} = l_t + \tau b_t \tag{4}$$

The 95% prediction interval could be obtained for y_{t+1} (when τ= 1), as shown in Equation (5):

$$(l_t + b_t) \pm z_{.025} s \tag{5}$$

From Equation (5), the 95% prediction interval for the τ-step forecast of $y_{t+\tau}$ could be obtained, as shown in Equation (6):

$$(l_t + \tau b_\tau) \pm z_{.025} s \times \sqrt{1 + \sum_{j=1}^{\tau-1} \alpha^2 (1 + j\tau)^2} \tag{6}$$

Where the standard deviation is obtained, as shown in Equation (7):

$$s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{j=1}^{t} [(y_j - (l_{j-1} + b_{j-1})]^2}{t-2}} \tag{7}$$

Measures of Incident Rate Forecast Error

It should be noted that every instance of the incident rate forecast has a random component, which is considered as the forecast error. Proper estimation of incident rate forecast errors would enable the management to determine whether the forecasting method accurately predicts incident rates. The accuracy of the incident rate forecast is necessary in order to effectively plan ahead, based on the need to adequately allocate resources to the safety intervention activities that minimize incident rates. The estimation of the incident rate forecast could also be used to determine whether the forecasting technique adopted is consistently producing very positive error (overestimation) or negative error (underestimation).

Forecasting errors that are within historical error estimates indicate a high level of consistency and accuracy of the forecasting technique. Forecast errors that are well beyond the historical estimates could indicate that the forecasting method is no longer accurate or appropriate. The commonly measured types of forecasting errors include the forecast or residual error, forecast accuracy, the mean squared error (MSE), absolute deviation (AD), mean absolute deviation (MAD), the mean absolute percentage error (MAPE), the percent mean absolute deviation (PMAD), bias (BIAS), and tracking signal (TS).

Forecast/Residual Error

The forecast or residual error (in terms of incident rate) is the difference between the forecast for period (t) and the actual incident rate in period (t). It may be important for the safety personnel to estimate the forecast error of incident rates far in advance. This is necessary in order to make decisions and planning on the type of resources to be allocated to a particular safety intervention activity before the occurrence of any incident. The forecast error (E_t) is measured as shown in Equation (8):

$$\mathbf{E_t = F_t - Y_t} \tag{8}$$

Where F_t is the forecast at period (t) and Y_t is the actual incident rate at period (t).

The percent forecast error is measured as the ratio of the forecast error and the forecast at period (t).The mathematical representation of the error (%) is shown in Equation (9):

$$\textbf{Error (\%)} = \frac{/Y_t - F_t/}{Y_t} \Rightarrow \frac{/Actual_t - Forecast_t/}{Actual_t} \tag{9}$$

The forecast error could be larger than the forecast or the actual incident rate at time (t), but not both. A zero forecast accuracy or very inaccurate forecast is obtained in situations where the percent forecast error [Error (%)] is higher than 100%.

Forecast Accuracy

Forecast accuracy is described as the proportion of deviation of the actual incident rate from the predicted incident rate in a period (t). Decreasing errors indicate increasing accuracy since the intention of the forecast is to predict incident rates that are identical to the actual values (Chopra and Meindl, 2006). Forecast accuracy is measured as shown in Equation (10).

$$\textbf{Accuracy (\%)} = \textbf{[1 – Error (\%)]} \tag{10}$$

In most situations, several forecasts yield accuracies which are between 0 and 100%. A perfect forecast is achieved when accuracy is 100%, while an absolutely incorrect or inaccurate forecast is obtained when accuracy is 0%.

Mean Squared Error

The mean squared error (MSE) is related to the variance of the forecast error. In this case, the random component of the incident rate is assumed to have a mean of zero and a variance of MSE. The mean squared error (variance of forecast error) is shown in Equation (11) as:

$$MSE = \frac{\sum_{t=1}^{N} E_t^2}{N} \tag{11}$$

Where N = Total number of observations or sample size.

Absolute Deviation

The absolute deviation (A_t) of an incident rate forecast in period (t) is defined as the absolute value of the forecast error (E_t) in period (t). The mathematical representation of the absolute deviation of the incident rate forecast is shown in Equation (12) as:

$$\boldsymbol{A_t} = |\, \mathbf{E_t} \,| \tag{12}$$

Mean Absolute Deviation

The mean absolute deviation (MAD) or the mean absolute error (MAE) is described as the average of the absolute deviation of the incident rate forecast for the total observations (Chopra and Meindl, 2006). The mean absolute deviation is expressed mathematically, as shown in Equation (13):

$$MAD = \frac{\sum_{t=1}^{N} / E_t /}{N} \tag{13}$$

In situations where the random component of the incident rate forecast is normally distributed, then the estimate of the standard deviation (σ), with the mean equal zero, could be obtained, as shown in Equation (14):

$$\sigma = 1.25\ \mathrm{MAD} \tag{14}$$

Mean Absolute Percentage Error

The mean absolute percentage error (MAPE) is described as the average value obtained from the summation of the absolute values of all the percentage errors (Chopra and Meindl, 2006). The mathematical representation of the mean absolute percentage error is shown in Equation (15):

$$MAPE = \frac{\sum_{t=1}^{N} \left| \frac{E_t}{Y_t} \right|}{N} \tag{15}$$

Percent Mean Absolute Deviation

The percent mean absolute deviation (PMAD) could be described as the parameter obtained from the division of the absolute total errors and the absolute total of the predicted incident rates. This is mathematically represented as shown in Equation (16):

$$PMAD = \frac{\sum_{t=1}^{N} |E_t|}{\sum_{t=1}^{N} |Y_t|} \tag{16}$$

Bias

The errors of a truly random, non-biased forecast should fluctuate around zero, with the best straight line passing through zero (Chopra and Meindl, 2006). The bias of the incident rate forecast could be obtained by summing up all the forecast errors, as shown in Equation (17):

$$Bias_N = \sum_{t=1}^{N} E_t \tag{17}$$

Tracking Signal

The tracking signal (TS) could be described as the ratio of the bias and the mean absolute deviation (MAD). This is represented as shown in Equation (18):

$$TS_t = \frac{Bias_t}{MAD_t} \tag{18}$$

The range of any tracking signal (TS) is $\pm$ 6. A forecast is either underestimated/ biased (TS < -6) or overestimated (TS > +6).

Data Collection Methodology

In order to show that incident rates could be effectively predicted, a safety intervention model, which was developed from an earlier safety and health intervention effectiveness research conducted by Oyewole and Haight (2009), was validated using the proposed forecasting methodology. An additional 52 weeks (one year) of data were collected from the same oil exploration and production company in the Niger Delta region of Nigeria. The collected data were based on the recommendations and the application of the suggested desirable resource allocation method proposed by Oyewole and Haight (2009). Using the averages of the past 52 weeks of the model development data, Holt's double exponential smoothing method was used to predict the incident rates for the next 52 weeks. Setting each of the alpha (α) and gamma (γ) levels to 0.25, the Holt's double exponential smoothing technique was used to predict the incident rate as shown in the Figure 1.

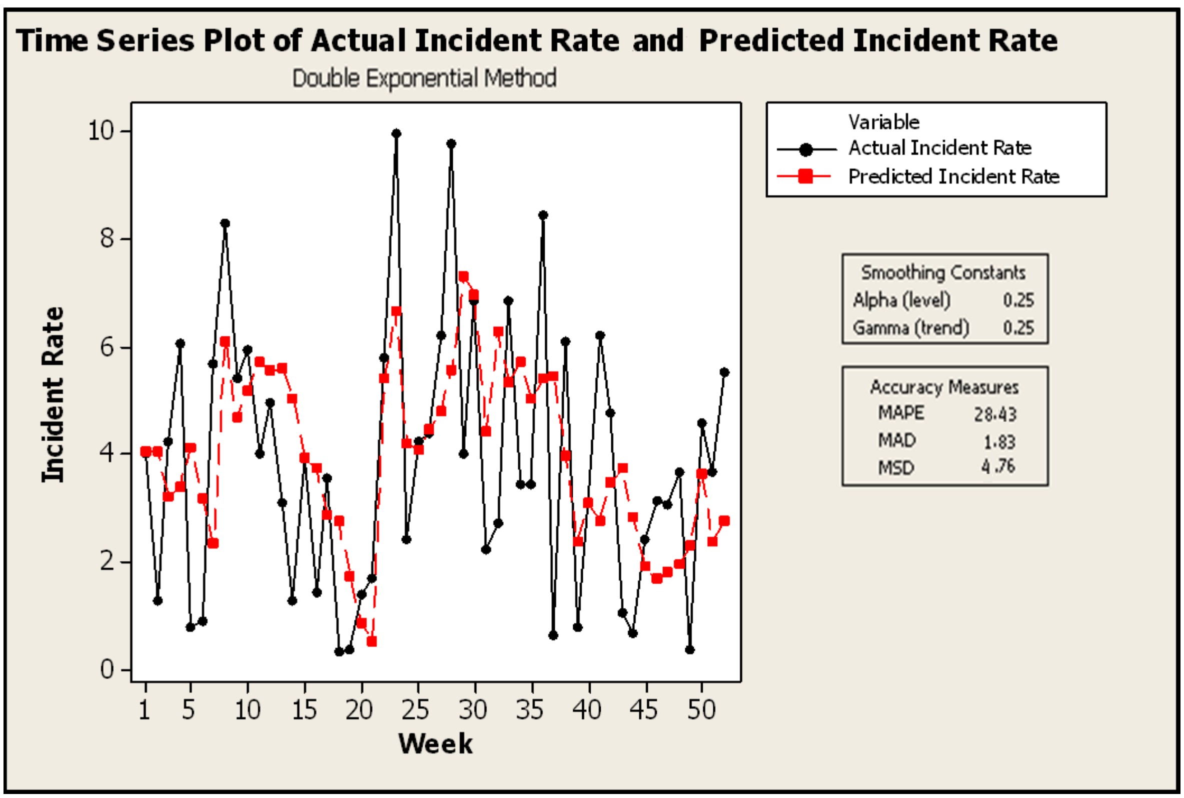

Figure 1. Time series plot for incident rate (predicted vs. actual)

The predicted incident rates and the actual incident rates were compared using statistical measures and forecasting errors, such as the forecast or residual error, forecast accuracy, mean squared error (MSE), absolute deviation (AD), mean absolute deviation (MAD), mean absolute percentage error (MAPE), the percent mean absolute deviation (PMAD), bias (BIAS), and tracking signal (TS). The statistical measurements used for the comparison of the predicted incident rates and the actual incident rates (based on the developed model) include the mean, median, quartile, and standard deviation. The statistical comparison of the smoothed/predicted incident rates and the actual incident rates is shown in Table 1.

Table 1. Measures of Statistical Comparison for Incident Rates (Predicted vs. Actual)

Statistical Measure	Smoothed/Predicted Incident Rates	Incident Rates (Actual)
Mean	3.75	3.82
Median	3.83	3.65
Standard Deviation	1.69	2.45
1st Quartile (Q1)	2.49	1.49
3rd Quartile (Q3)	5.04	5.63
Sample Size	52	52

The boxplots and the normal probability plots for the predicted incident rates and the incident rates (model) obtained from the analysis are shown in Figures 2 and 3, respectively.

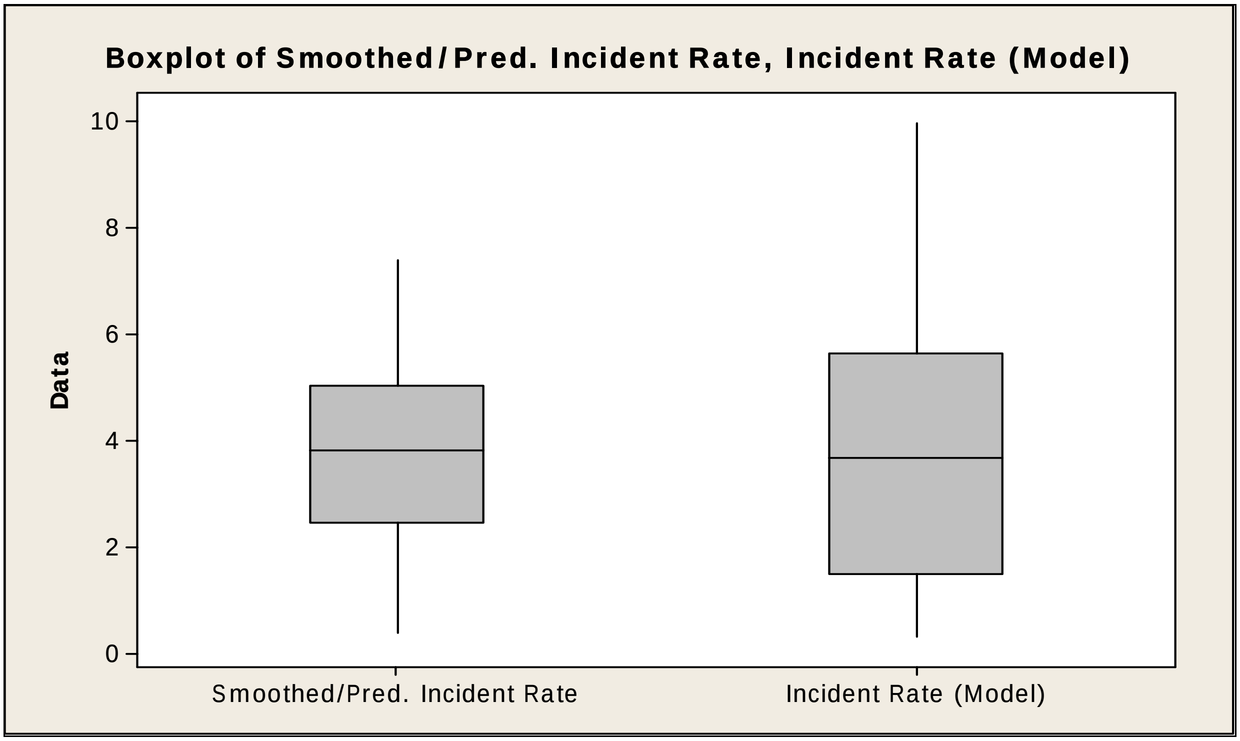

Figure 2. Boxplots for Predicted and Actual Incident Rates

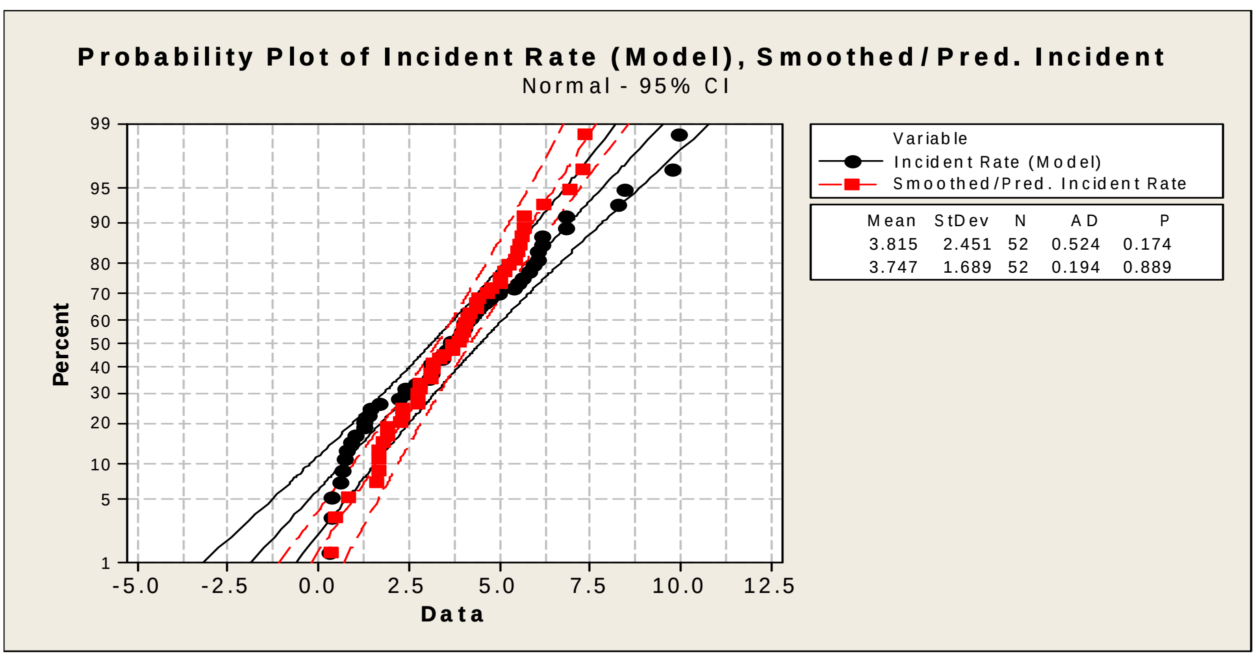

Figure 3. Normal Probability plots for Incident Rates (Predicted vs. Actual)

The histogram plots for the predicted incident rates and the incident rates (model) obtained from the analysis are shown in Figures 4 and 5, respectively.

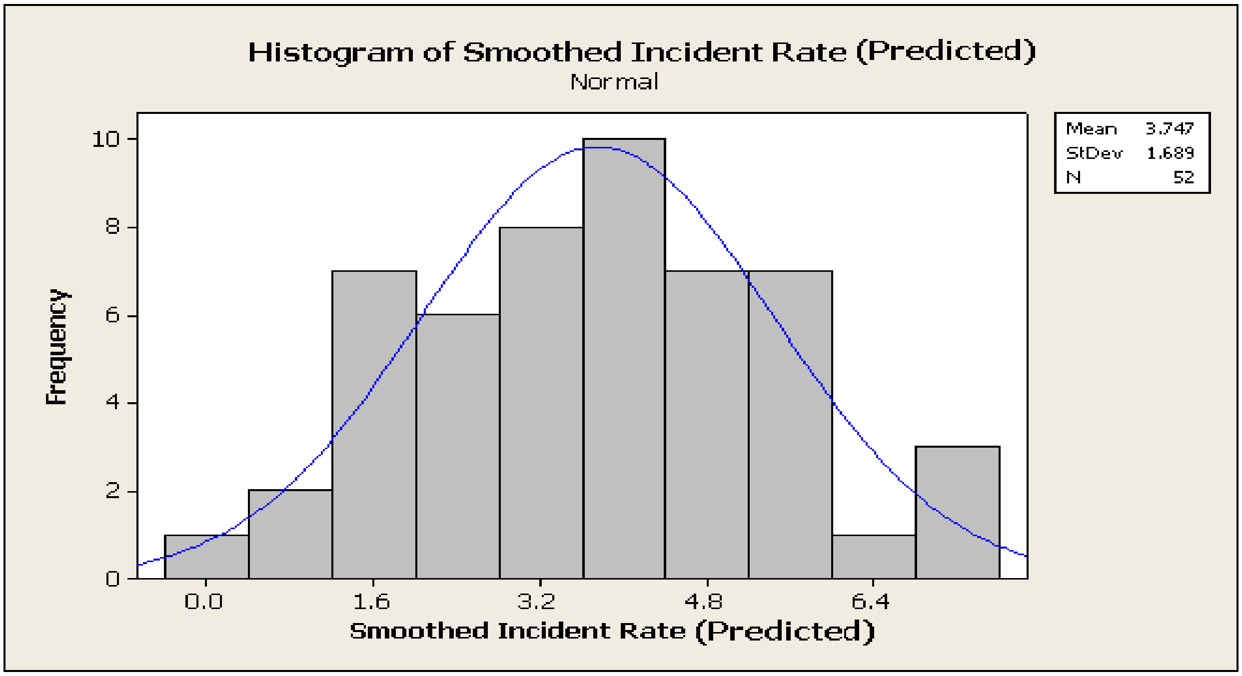

Figure 4. Histogram Plot for Predicted Incident Rates

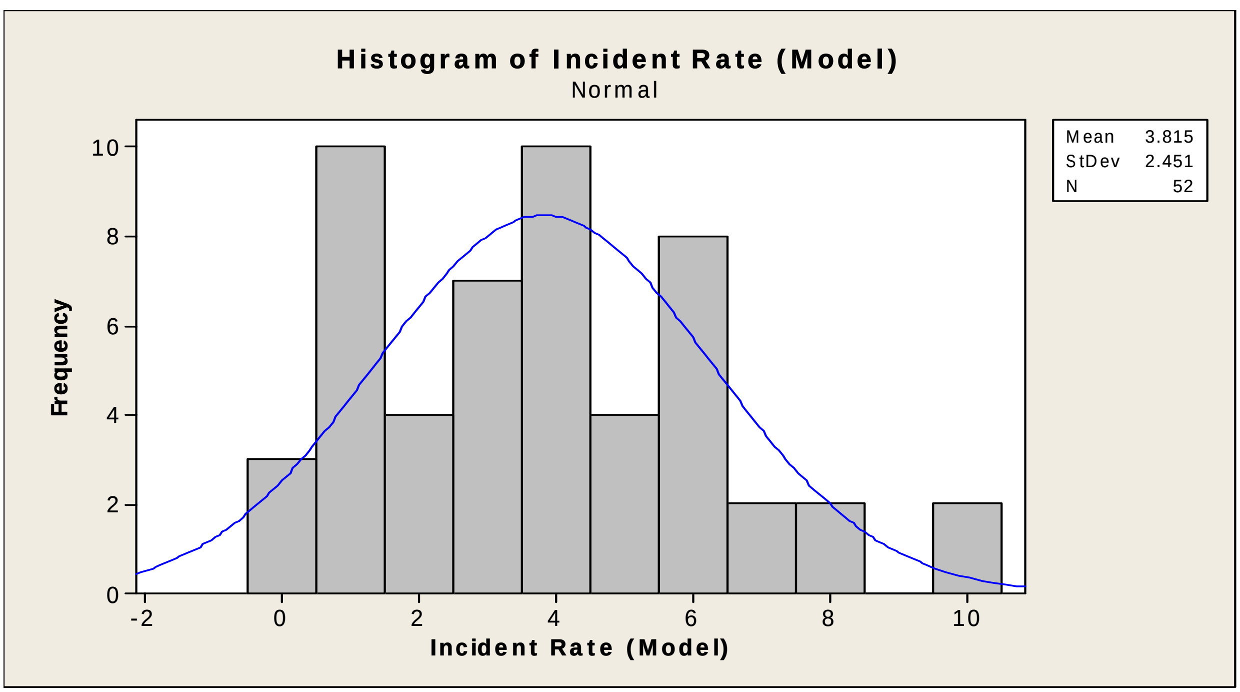

Figure 5. Histogram Plot for Actual Incident Rates

The model validation process, which was based on the measures of the forecasting errors, indicated that using the appropriate forecasting techniques, incident rates could be accurately predicted. Table 2 shows the values obtained from the analysis of the predicted and model incident rates, based on the forecast or residual error, forecast accuracy, mean squared error (MSE), absolute deviation (AD), mean absolute deviation (MAD), mean absolute percentage error (MAPE), the percent mean absolute deviation (PMAD), bias (BIAS), and tracking signal (TS).

Table 2. Measures of Forecasting Error for Predicted Incident Rates

Measure of Forecasting Error	Obtained Values
Mean Forecast/Residual Error	0.07
Forecast Accuracy (%)	71.58
Mean Squared Error (MSE)	4.76
Absolute Deviation (AD)	7.46
Mean Absolute Deviation (MAD)	1.83
Mean Absolute Percentage Error (MAPE)	28.43
Percent Mean Absolute Deviation (PMAD)	3.62
Bias (BIAS)	-7.46 (slope = 0.018)
Tracking Signal (TS)	-4.08

From Table 2, the obtained forecast accuracy value of 71.58% is within the acceptable limit of 0% to 100%. Based on the value of the obtained forecast accuracy, incident rates could be accurately predicted in more than 70% of the time. The obtained level of forecast accuracy (71.58%) could be explained due to the similarities in seasonality or trends. For example, periodic militant activities (sometimes occurring twice in a three-month period) could be a reason for the high spikes or outliers in the actual and the predicted incident rate data. An outlier could be described as an observation or data point which lies at an abnormal distance, when compared to the other values in the data.

Despite this observation, the obtained 71.58% incident rate prediction accuracy could be viewed as a starting advantage for health and safety managers and decision-makers, since it would enable the safety personnel to effectively plan ahead and appropriately allocate resources to the safety intervention activities, which would further lower incident rates. A bias value of -7.46 (with a slope value of 0.018) indicates that the forecast error is truly random and not biased in any way. Figure 6 shows the plot of the best straight line, indicating a slope value of +0.0184 for the linear equation of the forecast error. The slope value is not significant and revolves around zero.

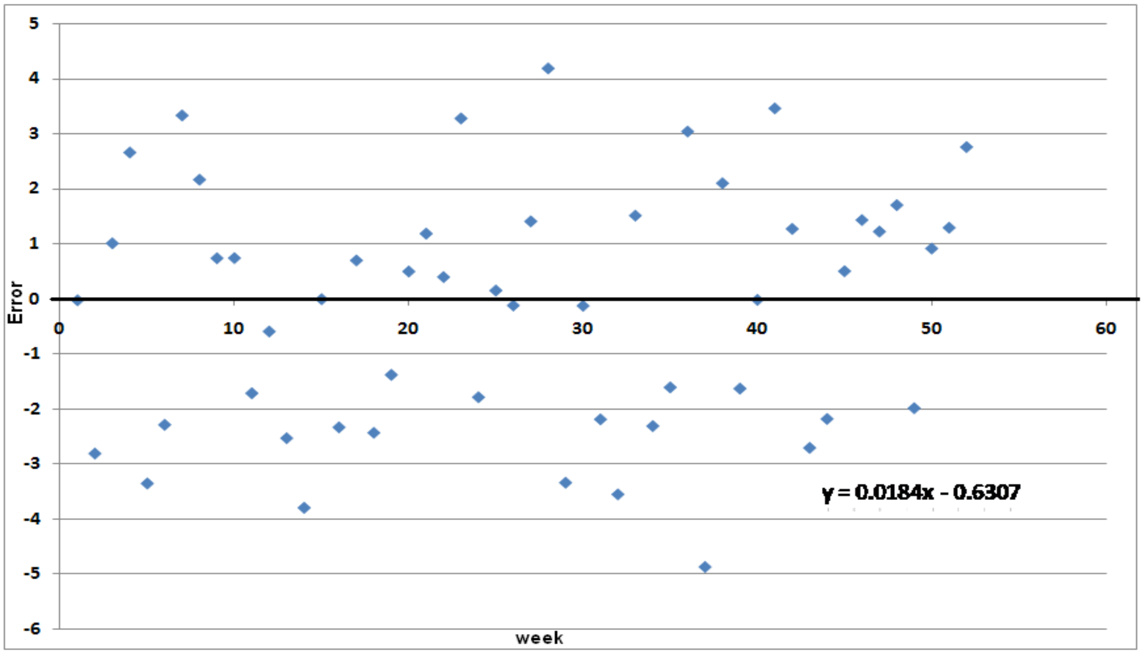

Figure 6. Residual Plot of Forecast Error

Discussion and Conclusion

The obtained value of the tracking signal (-4.08) shows that the signal for the forecast is not biased and within the acceptable range (TS $\pm$ 6). The negative sign of the tracking signal could be explained, based on the use of the averaging technique for the initial estimates of the incident rate and level. An excessively large negative value of the tracking signal could indicate the underestimation of the incident rate, while an excessively large positive value of the tracking signal could indicate the overestimation of the incident rate. This is not the case in this study, since the obtained value of the tracking signal is within the acceptable range.

The use of the double exponential smoothing technique (Holt's model) to predict incident rates provided the opportunity to validate the developed safety intervention model, based on the analysis of the measures of the forecasting errors. The obtained prediction accuracy (71.58%) could be improved upon in situations where sensitivity analysis is incorporated into the double exponential smoothing technique. Sensitivity analysis could be performed on the forecasting methodology in order to further reduce bias, improve the value of the tracking signal, and ultimately increase the level of the forecast accuracy. The adjustment of the alpha (α) and gamma (γ) levels, based on the preference of the safety personnel, as well as the selection of a better estimation technique for the initial value of the incident rate and trend level, could also improve the prediction capability and accuracy of the forecasting methodology.

This research shows that injury incident rates can be effectively predicted, using traditional business decision-making and forecasting tools and techniques. The ability to predict incident rates helps in the identification of significant safety intervention factors, which are aimed at reducing safety intervention costs. This would also make it possible for managers and top-level executives to better understand the financial implications and consequences of ineffective safety programs and policies. This could be seen as an added advantage, since the obtained information from the incident rate forecast could be used to make better decisions on the level of resources to allocate to safety and health programs.

Future Work

The expansion of the current sample size of the data by incorporating and comparing safety activities from other units or organizations would be beneficial in order to achieve higher prediction accuracies for incident rates. Increasing the data sample size would also allow management to adequately understand the impact of allocating sufficient budget and resources to the various tasks, operations, intervention activities, and safety programs that reduce incident rates. Setting of safety decision-making standards by incorporating weights to the factors would also be beneficial to this study. This could provide a more realistic value of incident rates and could indicate the level of willingness of the management in the allocation of resources towards the safety activities with the help of an effective incident rate prediction technique.

References

Armstrong, J.S. 2001. *Principles of Forecasting: A Handbook for Researchers and Practitioners* (International Series in Operations Research & Management Science). Norwell, MA: Kluwer Academic Publishers.

Bowerman, B.J. and R.T. O'Connell. 1993. *Forecasting and Time Series: An Applied Approach.* Belmont, CA: Duxbury Thomson Learning.

Chopra, S and P. Meindl. 2006. *Supply Chain Management: Strategy, Planning and Operation,* 3rd edition. Upper Saddle River, NJ: Prentice Hall.

Haight, J. M., R. E. Thomas, L. A. Smith, R. L. Bulfin Jr. and B. K. Hopkins. 2001a. "Intervention Effectiveness Research: Phase 1, Developing a Mathematical Relationship between Interventions and Incident Rates for the Design of a Loss Prevention System." American Society of Safety Engineers Technical Forum: 38-44.

Haight, J. M., R. E. Thomas, L. A. Smith, R. L. Bulfin Jr. and B. K. Hopkins. 2001b. "Intervention Effectiveness Research: Phase 2 Design, Optimization & Verification of the Loss Prevention System & Analysis Models." *Professional Safety: The Journal of the America Society of Safety Engineers:* 33-37.

Iyer, P., J.M. Haight, E. del Castillo, B.W. Tink, and P.W. Hawkins. 2004. "Intervention effectiveness research: Understanding and optimizing industrial safety programs using leading indicators." *Chemical Health and Safety*, 11(2), 9-19.

Iyer, P., J.M. Haight, E. del Castillo, B.W. Tink, and P.W. Hawkins. 2005. "A Research Model—Forecasting Incident Rates from Optimized Safety Program Intervention Strategies." *Journal of Safety Research,* Vol. 36, Number. 4, pp. 341-351.

Oyewole, S.A and J.M. Haight. 2009. "Making the Business Case: Assessment of Safety Intervention and Resource Allocation and Optimization." In the Proceedings of 2009 Safety Conference and Exposition organized by the American Society of Safety Engineers (ASSE), San Antonio, TX. June 28–July 1, 2009.

Toffel, M. W. and L. R. Birkner. 2002. "Estimating and Controlling Workplace Risk: An Approach for Occupational Hygiene and Safety Professionals." *Applied Occupational and Environmental Hygiene,* Volume 17, Number 7, pp. 477-485(9).